The tiling by the minimal separators of a junction tree and applications to graphical models

Durham, 2008

*

*G. Letac, Université Paul Sabatier, Toulouse. Joint work with H. Massam

Motivation : the Wishart distributions on decomposable graphs.

We denote by $\mathcal{S}_n$ the space of symmetric real matrices of order n and by $P_n \subset \mathcal{S}_n$ the cone of positive definite matrices. Let $G = (V, \mathcal{E})$ be a decomposable graph with $V = \{1, \dots, n\}$. The subspace $ZS_G \subset \mathcal{S}_n$ is the space of symmetric matrices (s_{ij}) with zeros prescribed by G, that means $s_{ij} = 0$ when $\{i, j\} \notin \mathcal{E}$. We denote

$$P_G = ZS_G \cap \mathcal{P}_n.$$

A space isomorphic to ZS_G is the space IS_G of symmetric incomplete matrices which are actually real functions on the union of the set V and the set $\mathcal{E}$ of edges.

We denote by π the natural projection of $\mathcal{S}_n$ on IS_G and denote $Q_G = \pi(P_G^{-1})$. Three equivalent properties

1. $Q_G = \pi(P_G^{-1})$ (definition)
2. Q_G is the open convex cone which is the dual of the cone P_G.
3. the restriction x_C of $x \in Q_G$ to any clique C is positive definite (Hélène's definition).

Example : If

$$G = \bullet 1 - \bullet 2 - \bullet 3$$

the cone P_G is the set of positive definite matrices of the form

$$\begin{bmatrix} y_1 & y_{12} & 0 \\ y_{12} & y_2 & y_{23} \\ 0 & y_{32} & y_3 \end{bmatrix}$$

The cone Q_G is the set of incomplete matrices of the form

$$\begin{bmatrix} x_1 & x_{12} & \\ x_{12} & x_2 & x_{23} \\ & x_{32} & x_3 \end{bmatrix}$$

such that the two submatrices associated to the two cliques

$$\begin{bmatrix} x_1 & x_{12} \\ x_{12} & x_2 \end{bmatrix}, \quad \begin{bmatrix} x_2 & x_{23} \\ x_{32} & x_3 \end{bmatrix}$$

are positive definite.

The bijection between P_G and Q_G. Let G decomposable, let $\mathcal{C}$ and $\mathcal{S}$ be the families of cliques and minimal separators. If $x \in Q_G$ define the Lauritzen function :

$$y = \psi(x) = \sum_{C \in \mathcal{C}} [x_C^{-1}]_0 - \sum_{S \in \mathcal{S}} \nu(S)[x_S^{-1}]_0$$

where $[a]_0$ means 'extension by zeros' of a principal submatrix a of $\mathcal{S}_n$ and where $\nu(S)$ is a certain positive integer called multiplicity of S.

Theorem 1. The map

$$x \mapsto y = \psi(x)$$

is a diffeomorphism from Q_G onto P_G. Its inverse $y \mapsto x$ from P_G onto Q_G is $x = \pi(y^{-1})$.

Let us fix $\alpha : \mathcal{C} \to \mathbb{R}$ and $\beta : \mathcal{S} \to \mathbb{R}$ and let us introduce the function $x \mapsto H(\alpha, \beta; x)$ on Q_G by

$$H(\alpha, \beta; x) = \frac{\prod_{C \in \mathcal{C}} \det(x_C)^{\alpha(C)}}{\prod_{S \in \mathcal{S}} \det(x_S)^{\nu(S)\beta(S)}}.$$

Define the measure on Q_G by

$$\mu_G(dx) = H(-\frac{1}{2}(|C|+1), -\frac{1}{2}(|S|+1; x)\mathbf{1}_{Q_G}(x)dx.$$

Perfect orderings of the cliques. Let $\mathcal{C}$ be the family of the k cliques of the connected graph (not necessarily decomposable). Consider a bijection $P : \{1, \ldots, k\} \to \mathcal{C}$ and

$$S_P(j) = [P(1) \cup P(2) \cup \ldots \cup P(j-1)] \cap P(j)$$

for $j \geq 2$. Then the ordering P is said to be *perfect* if there exists $i_j < j$ such that

$$S_P(j) \subset P(i_j).$$

This is a deep notion : a connected graph is decomposable if and only if a perfect ordering of the cliques exists. Furthermore if G is decomposable and if P is perfect then $S_P(j)$ is a minimal separator.

Let us fix a perfect ordering P of the set $\mathcal{C}$ of the cliques. For a fixed minimal separator S consider the set of cliques $J(P,S) =$

$\{C \in \mathcal{C}\ ;\ \exists j \geq 2$ such that $P(j) = C$ et $S_P(j) = S\}$.

An important result is that if P is a perfect ordering and if for all $S \in \mathcal{S}$ different from $S_P(2)$ one has

$$\sum_{C \in J(P,S)} (\alpha(C) - \beta(S)) = 0$$

(we denote by $\mathcal{A}_P$ this set of (α, β)'s) then by a long calculation one sees that there exists a number $\Gamma(\alpha, \beta)$ with the following eigenvalue property : for all $y \in P_G$

$$\int_{Q_G} e^{-\operatorname{tr} xy} H(\alpha, \beta; x) \mu_G(dx) = \Gamma(\alpha, \beta) H(\alpha, \beta; \pi(y^{-1})).$$

(L.-Massam, *Ann. Statist.* 2007).

A reformulation is

$$\int_{Q_G} e^{-\operatorname{tr} x\psi(x_1)} H(\alpha,\beta;x)\mu_G(dx) = \Gamma(\alpha,\beta)H(\alpha,\beta;x_1)$$

namely the functions $x \mapsto H(\alpha,\beta;x)$ are eigenfunctions of the operator $f \mapsto K(f)$ on functions on Q_G defined by

$$K(f)(x_1) = \int_{Q_G} e^{-\operatorname{tr} x\psi(x_1)} f(x)\mu_G(dx).$$

This leads to the definition of the Wishart distributions on Q_G by

$$\frac{1}{\Gamma(\alpha,\beta)H(\alpha,\beta;\pi(y^{-1}))}e^{-\operatorname{tr}(xy)}H(\alpha,\beta;x)\mu_G(dx)$$

They are therefore indexed by the shape parameters (α,β) and by the scale parameter $y \in P_G$.

There is an other family of similar formulas where the roles of P_G and Q_G are exchanged that I have no time to describe.

Homogeneous graphs and the graph A_4. I have to mention that if G is homogeneous, that is if P_G is an homogeneous cone (This happens if and only if G does not contains the chain

$$A_4 : \bullet - \bullet - \bullet - \bullet$$

as an induced graph), the above formulas hold for a wider range of parameters α and β than the union of $\mathcal{A}_P$ where P runs all the perfect orderings. Thus the simplest non homogeneous graph is $G = A_4 = \bullet 1 - \bullet 2 - \bullet 3 - \bullet 4$ with cliques and separators

$$C_1 = \{1, 2\},\ C_2 = \{2, 3\},\ C_3 = \{3, 4\},$$

$$S_2 = \{2\},\ S_3 = \{3\}.$$

An element of Q_G has the form

$$x = \begin{bmatrix} x_1 & x_{12} & & \\ x_{21} & x_2 & x_{23} & \\ & x_{32} & x_3 & x_{34} \\ & & x_{43} & x_4 \end{bmatrix}$$

for $x \in Q_G$, with $x_{ij} = x_{ji}$,

Let $\alpha_i = \alpha(C_i), i = 1, 2, 3 \ \ \beta_i = \beta(S_i), i = 2, 3$.
Define $\mathcal{D} =$

$$\{(\alpha, \beta) \mid \alpha_i > \frac{1}{2},\ i = 1, 2, 3,\ \alpha_1 + \alpha_2 > \beta_2, \alpha_2 + \alpha_3 > \beta_3\}.$$

Then the following integral (a 7-uple integral!) converges for all $\sigma \in Q_{A_4}$ if and only if (α, β) is in $\mathcal{D}$. Under these conditions, it is equal to

$$\int_{Q_G} e^{-\langle x, \psi(\sigma)\rangle} H_G(\alpha, \beta; x) \mu_G(dx)$$

$$= \frac{\Gamma(\alpha_1 - \frac{1}{2})\Gamma(\alpha_2 - \frac{1}{2})\Gamma(\alpha_3 - \frac{1}{2})}{\Gamma(\alpha_2)}$$
$$\times \Gamma(\alpha_1 + \alpha_2 - \beta_2)\Gamma(\alpha_2 + \alpha_3 - \beta_3)$$
$$\times \pi^{\frac{3}{2}} \sigma_{1\cdot 2}^{\alpha_1} \sigma_{2\cdot 3}^{\alpha_1 + \alpha_2 - \beta_2} \sigma_{3\cdot 2}^{\alpha_2 + \alpha_3 - \beta_3} \sigma_{4\cdot 3}^{\alpha_3}$$
$$\times {}_2F_1(\alpha_1 + \alpha_2 - \beta_2, \alpha_2 + \alpha_3 - \beta_3, \alpha_2, \frac{\sigma_{23}^2}{\sigma_2 \sigma_3})$$

where ${}_2F_1$ denotes the hypergeometric function.

NB $\sigma_{i.j}$ means $\sigma_i - \sigma_{ij}\sigma_j^{-1}\sigma_{ji}$, thus line 3 is a function of type $H(\alpha, \beta; \sigma)$.

The two lessons of the example

$$A_4 : \bullet 1 - \bullet 2 - \bullet 3 - \bullet 4$$

1. The integral has the form $CH(\alpha,\beta;\sigma)$ if and only if the hypergeometric function degenerates (we mean when $c = a$ or b for ${}_2F_1(a,b;c;x)$. Therefore (α,β) satisfies the eigenvalue property if and only if it is in the union of the $\mathcal{A}_P$'s for the 4 perfect orderings P of A_4.
2. The 4 perfect orderings are

$$P_1 = C_1C_2C_3,\ P_2 = C_2C_1C_3,$$

$$P_3 = C_3C_2C_1,\ P_4 = C_2C_3C_1$$

but

$$\mathcal{A}_{P_1} = \mathcal{A}_{P_2} = \{\alpha_2 = \beta_2\} \cap \mathcal{D}$$

$$\mathcal{A}_{P_3} = \mathcal{A}_{P_4} = \{\alpha_3 = \beta_3\} \cap \mathcal{D}$$

Why ? As we are going to see, this is because P_1 and P_2 share the same initial minimal separator, as well as P_4 and P_3.

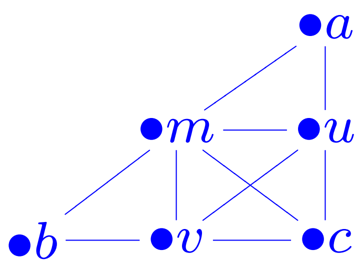
What one needs to review about decomposable graphs

1. Junction trees.
2. Minimal separators
3. The two definitions of the multiplicity of a minimal separator.

Junction trees The cliques of a graph are its maximal complete subsets. A junction tree has the set of cliques as set of vertices and is such that if the clique C'' is on the unique path from C to C' then $C'' \supset C \cap C'$. For instance

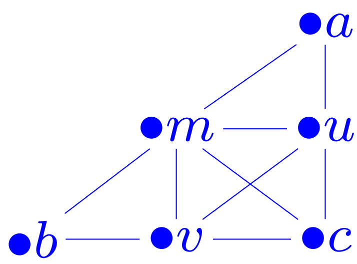
$$\bullet 1 - \bullet 2 - \bullet 3$$

is a junction tree for the decomposable graph



where the three cliques are $1 = (amu)$, $2 = (muvc)$ and $3 = (bmv)$. A connected graph is decomposable if and only if a junction tree exists (a neat proof of this is given by Blair and Peyton in 1991)

Minimal separators If a and b are not neighbors $S \subset V$ is a separator of a and b if any path from a to b hits S



For instance $muvc$ is a separator of a and b. If nothing can be taken out, S is a *minimal* separator of a and b. Finally S is *minimal* separator by itself if there exist non adjacent a and b such that S is a minimal separator of a and b. There are not so many of them, strictly less that the number of cliques anyway. They are mu and mv in the example. A connected graph is decomposable if and only if all the minimal separators are complete (Dirac 1961).

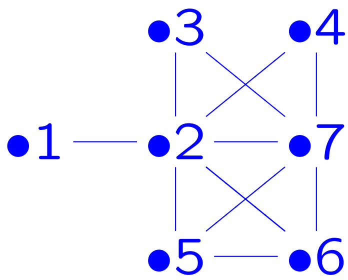
Topological multiplicity of a minimal separator Let S be a minimal separator of a decomposable graph $(V, \mathcal{E})$. Let $\{V_1, \ldots, V_p\}$ be the connected components of $V \setminus S$ (of course $p \geq 2$). Let q be the number of $j = 1, \ldots, p$ such that S is *NOT* a clique of $S \cup V_j$. The number $\nu(S) = q - 1$ is called the topological multiplicity of S.

Multiplicity of a minimal separator from a perfect ordering If P is a perfect ordering and if S is a minimal separator, denote by $\nu_P(S)$ the number of $j = 2, \ldots, k$ such that $S = S_P(j)$; recall

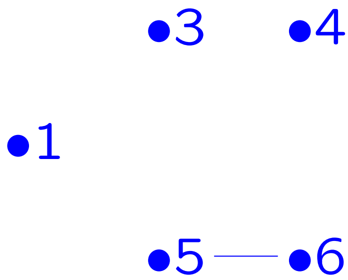
$$S_P(j) = [P(1) \cup P(2) \cup \ldots \cup P(j-1)] \cap P(j).$$

(The topological multiplicity is introduced by Lauritzen, Speed and Vivayan in 1984). Question : one observes that in all cases the two definitions of multiplicity coincide. Why ? Answer later on.

Example : If I remove the minimal separator $S = 27$ to its graph

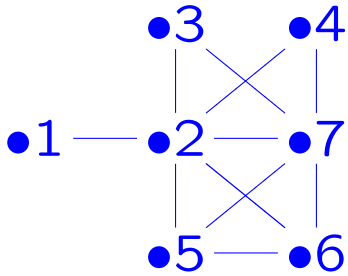


four connected components are obtained :



If I add S to each of them, thus for component 1 I obtain the graph

for which $S = 27$ is a clique. This is not the case for the three other connected components 3, 4 et 56. Therefore $q = 3$, the topological multiplicity of 27 is 2.



Similarly since the cliques are $C_1 = 12$, $C_2 = 237$, $C_3 = 247$, $C_4 = 2567$ one can see that $P = C_1C_2C_3C_4$ is perfect that $S_P(3) = S_P(4) = 27$ and that $\nu_P(27) = 2 = \nu(P)$.

Question :

Why do we have always $\nu_P(S) = \nu(S)$?

Tiling of a junction tree by the minimal separators If $(H, \mathcal{E}(H))$ is a tree (undirected) with vertex set H and edge set $\mathcal{E}(H)$ a *tiling* of H is a family $\mathcal{T}$ of subtrees

$$\mathcal{T} = \{T_1, \ldots, T_p\}$$

of H such that if $\mathcal{E}(T_i)$ is the edge set of T_i then

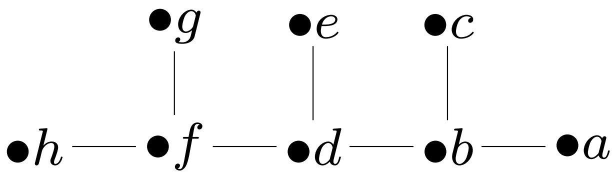
$$\{\mathcal{E}(T_1), \ldots, \mathcal{E}(T_q)\}$$

is a partition of $\mathcal{E}(H)$. This implies

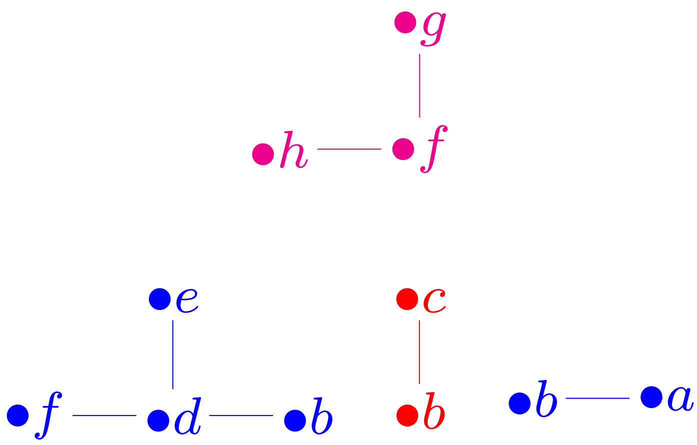
$$T_1 \cup \ldots \cup T_q = H$$

although $(T_1, \ldots, T_p)$ is not a partition of the set H.

Example



the tiles of the tiling can be chosen as



Theorem 1.

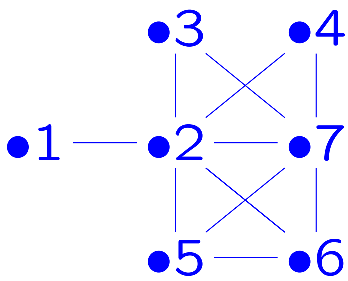
Let $G = (V, \mathcal{E})$ be a decomposable graph and let $(\mathcal{C}, \mathcal{E}(\mathcal{C}))$ be a junction tree of G. Let $\mathcal{S}$ be the family of minimal separators of G. There exists a unique tiling $\mathcal{T}$ of the tree $(\mathcal{C}, \mathcal{E}(\mathcal{C}))$ by subtrees and a bijection $S \mapsto T_S$ from $\mathcal{S}$ towards $\mathcal{T}$ with the following property : for all $S \in \mathcal{S}$ the edges of T_S are the edges $\{C, C'\}$ such that $S = C \cap C'$.

Under these circumstances the number of edges of T_S is the topological multiplicity of S. Furthermore if C and C' are two distinct cliques consider the unique path $(C = C_0, C_1, \ldots, C_q = C')$ from C to C'. Let $S_i \in \mathcal{S}$ such that $\{C_{i-1}, C_i\}$ is in T_{S_i}. Then

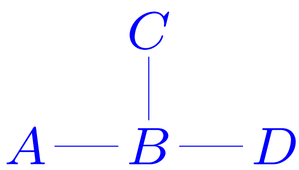
$$C \cap C' = \cap_{i=1}^{q} S_i.$$

In particular $C \cap C' = S$ if C and C' are in T_S.

Consider again the example :



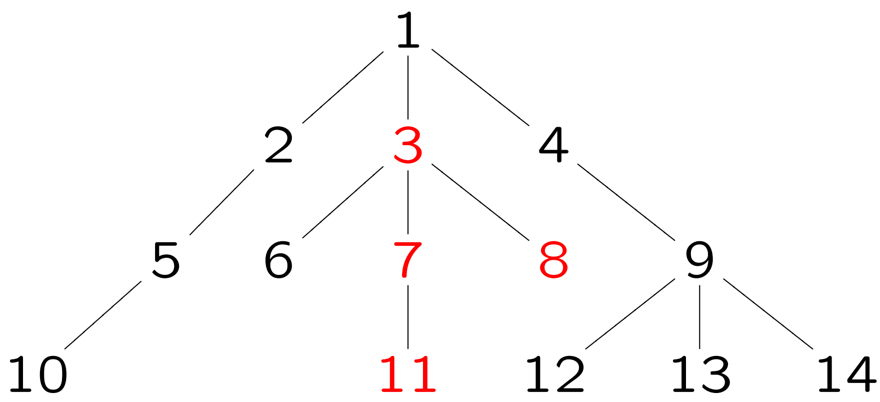
There are 4 cliques $A = \{1, 2\}$, $B = \{2, 3, 7\}$, $C = \{2, 4, 7\}$, $D = \{2, 5, 6, 7\}$ and two minimal separators $U = \{2\}$, $V = \{2, 7\}$. The ordering $ABCD$ of the cliques is perfect with $S_2 = U$ et $S_3 = S_4 = V$. Thus V has multiplicity 2 and U has multiplicity 1. Consider the junction tree



Then $T_U = AB$ et $T_V = BCD$.

Junction trees and perfect orderings of cliques. Recall that saying that P is a perfect ordering of the set $\mathcal{C}$ of the k cliques of a decomposable graph is to say that there exists $i_j < j$ such that $S_P(j) \subset P(i_j)$. There exist in general several possible i_j's. Actually we fix one such i_j for each j and we create the graph having $\mathcal{C}$ as vertex set with having the $k-1$ edges $\{P(i_j), P(j)\}$. A beautiful result of Beeri, Fagin, Maier and Yannakakis (1983) claims that this graph is a junction tree and conversely that any junction tree can be constructed from a perfect ordering and from a choice of the $j \mapsto i_j$. Let us say that a junction tree is *adapted to* the perfect ordering P if there exists a choice $j \mapsto i_j$ giving the tree.

Tiling by minimal separators and perfect orderings of the cliques. Let P be a perfect ordering of the set $\mathcal{C}$ of the k cliques of a decomposable graph Consider now a junction tree adapted to P and let $\mathcal{T}$ be the tiling of this tree by the minimal separators. We transform this undirected tree into a rooted tree by taking $P(1)$ as a root. This transforms $\mathcal{C}$ into a partially ordered set : $C \preceq C'$ if the unique path from $P(1)$ to C' passes through C.



Let S be in the set $\mathcal{S}$ of the minimal separators. Recall that we have considered before the set of cliques $J(P, S) =$

$$\{C \in \mathcal{C} \ ; \ \exists j \geq 2 \text{ such that } P(j) = C \text{ et } S_P(j) = S\}.$$

Just remark that $\nu_P(S) = |J(P, S)|$. Now for all $S \in \mathcal{S}$ the subtree T_S has a *minimal point* $M(S)$ for this partial order. Here is now a useful result ruling out the old contest between multiplicities (recall that the number of vertices of a tree is the number of edges plus one) :

Theorem 2.

$J(P, S) = T_S \setminus \{M(S)\}$. In particular $\nu_P(S)$ is the topological multiplicity $|T_S| - 1$ of S.

Actually $J(P,S)$ depends on S and on $S_P(2)$ only :

Theorem 3. Let P and P' two perfect orderings such that $P(1)\cap P(2) = P'(1)\cap P'(2)$, that is to say $S_P(2) = S_{P'}(2)$ (denoted S_2). Then $J(P,S) = J(P',S)$ if $S \neq S_2$ and

$$J(P,S_2)\cup\{P(1)\} = J(P',S_2)\cup\{P'(1)\}.$$

Conclusion : Consequences for the Wishart distributions on decomposable graphs.

Recall that given a perfect ordering P, the set $\mathcal{A}_P$ of acceptable shape parameters (α, β) for the Wishart distribution is the set of (α, β) such that for all minimal separators S we have

$$\sum_{C \in J(P,S)} (\alpha(C) - \beta(S)) = 0.$$

Thus this crucial set $\mathcal{A}_P$ depends entirely on the family of subsets of cliques

$$\mathcal{F}_P = \{J(P, S);\ S \in \mathcal{S}\}.$$

This tiling process has shown that actually the family $\mathcal{F}_P$ -and therefore the set $\mathcal{A}_P$ of shape parameters - depends only on the first minimal separator $S_P(2)$ of the perfect ordering P.